

1. HDP 中安装 Atlas
1. 1. 启动安装引导程序

1. 2. 选择服务

1. 3. 分配Master

1. 4. 分配Slaves and Clients

1. 5. 配置 Atlas 用户名和密码

1. 6. 部署

1. 7. 部署预览

1. 8. 等待部署完成
2. 验证
2. 1. 验证 Atlas

2. 2. 验证 HBase 表是否创建

2. 3. 验证 Solr 集合是否创建

2. 4. 验证 Kafka Topic 是否创建

1. HDP 中安装 Atlas

1.1. 启动安装引导程序



Service	Version	Status	Description
HDFS	3.1.1	Installed	Apache Hadoop Distributed File System
YARN	3.1.1	Installed	Apache Hadoop NextGen MapReduce (YARN)
MapReduce2	3.1.1	Installed	Apache Hadoop NextGen MapReduce (YARN)
Tez	0.9.1	Installed	Tez is the next generation Hadoop Query Processing framework written on
Hive	3.1.0	Installed	Data warehouse system for ad-hoc queries & analysis of large datasets and
HBase	2.0.2	Installed	Non-relational distributed database and centralized service for configuratic
Pig	0.16.0	Add Service	Scripting platform for analyzing large datasets
Sqoop	1.4.7	Add Service	Tool for transferring bulk data between Apache Hadoop and structured dat
Oozie	4.3.1	Add Service	System for workflow coordination and execution of Apache Hadoop jobs. T
ZooKeeper	3.4.6	Installed	Centralized service which provides highly reliable distributed coordination
Storm	1.2.1	Add Service	Apache Hadoop Stream processing framework
Accumulo	1.7.0	Add Service	Robust, scalable, high performance distributed key/value store.
Infra Solr	0.1.0	Installed	Core shared service used by Ambari managed components.
Ambari Metrics	0.1.0	Installed	A system for metrics collection that provides storage and retrieval capabili
Atlas	1.1.0	Add Service	Atlas Metadata and Governance platform
Kafka	2.0.0	Installed	A high-throughput distributed messaging system

1.2. 选择服务

Add Service Wizard

<input type="checkbox"/>	Oozie	4.3.1	System for workflow coordination and execution of Apache Hadoop which relies on and will install the ExtJS Library.
<input checked="" type="checkbox"/>	ZooKeeper	3.4.6	Centralized service which provides highly reliable distributed coordination.
<input type="checkbox"/>	Storm	1.2.1	Apache Hadoop Stream processing framework
<input type="checkbox"/>	Accumulo	1.7.0	Robust, scalable, high performance distributed key/value store.
<input checked="" type="checkbox"/>	Infra Solr	0.1.0	Core shared service used by Ambari managed components.
<input checked="" type="checkbox"/>	Ambari Metrics	0.1.0	A system for metrics collection that provides storage and retrieval.
<input checked="" type="checkbox"/>	Atlas	1.1.0	Atlas Metadata and Governance platform
<input checked="" type="checkbox"/>	Kafka	2.0.0	A high-throughput distributed messaging system
<input type="checkbox"/>	Knox	1.0.0	Provides a single point of authentication and access for Apache Hadoop.
<input type="checkbox"/>	Log Search	0.5.0	Log aggregation, analysis, and visualization for Ambari managed components.
<input type="checkbox"/>	Ranger	1.2.0	Comprehensive security for Hadoop
<input type="checkbox"/>	Ranger KMS	1.2.0	Key Management Server
<input checked="" type="checkbox"/>	SmartSense	1.5.1 2.7.4.0-118	SmartSense - Hortonworks SmartSense Tool (HST) helps quickly troubleshoot support cases and receive cluster-specific recommendations.
<input checked="" type="checkbox"/>	Spark2	2.3.2	Apache Spark 2.3 is a fast and general engine for large-scale data processing.

在分配Master环节，请按照规划选择Solr和Atlas部署的节点，可以部署多实例实现HA：

1.5. 配置 Atlas 用户名和密码

配置用户名和密码：admin 和 QWer_1234

Add Service Wizard

✓ Choose Services

✓ Assign Masters

✓ Assign Slaves and Clients

4 Customize Services

5 Review

6 Install, Start and Test

7 Summary

Customize Services

We have come up with recommended configurations for the services you selected. Customize them as you see fit.

HDFS YARN MAPREDUCE2 TEZ HIVE HBASE ZOOKEEPER INFRA SOLR AMBARI METRICS ATLAS KAFKA SMARTSENSE SPARK2 ZEPPELIN NOTEBOOK MISC

There are 5 configuration changes in 2 services Show Details

Group

Default (3)

AUTHENTICATION

ADVANCED

Advanced application-properties

Advanced atlas-env

Admin password

.....

.....

Admin username

admin

1.6. 部署

Add Service Wizard

Configurations

Highly Recommended Configurations 2

Please review the following recommended changes, and click on the property name to change its value.

Type	Service	Property	Current Value	Description
Warning	Ambari Metrics	hbase_master_heapsize	896	Value is less than the recommended default of 1152 HBase Master Heap Size. In embedded mode, total heap size is sum of master and regionserver heap sizes.
Warning	Atlas	atlas.graph.storage.hostname	bigdata22,bigdata23,bigdata24	Atlas is configured to use the HBase installed in this cluster. If you would like Atlas to use another HBase instance, please configure this property and HBASE_CONF_DIR variable in atlas-env appropriately.

CANCEL

PROCEED ANYWAY

1.7. 部署预览

Add Service Wizard

✓ Choose Services

✓ Assign Masters

✓ Assign Slaves and Clients

✓ Customize Services

5 Review

6 Install, Start and Test

7 Summary

Review

Please review the configuration before installation

Admin Name : admin

Cluster Name : my_test_hdp

Total Hosts : 3 (0 new)

Repositories:

redhat7 (HDP-3.1):
http://bigdata22/hdp314-centos7/HDP/centos7/3.1.4.0-315

redhat7 (HDP-3.1-GPL):
http://bigdata22/hdp314-centos7/HDP-GPL/centos7/3.1.4.0-315

redhat7 (HDP-UTILS-1.1.0.22):
http://bigdata22/hdp314-centos7/HDP-UTILS/centos7/1.1.0.22

Services:

Atlas

Metadata Server : bigdata23

Zeppelin Notebook

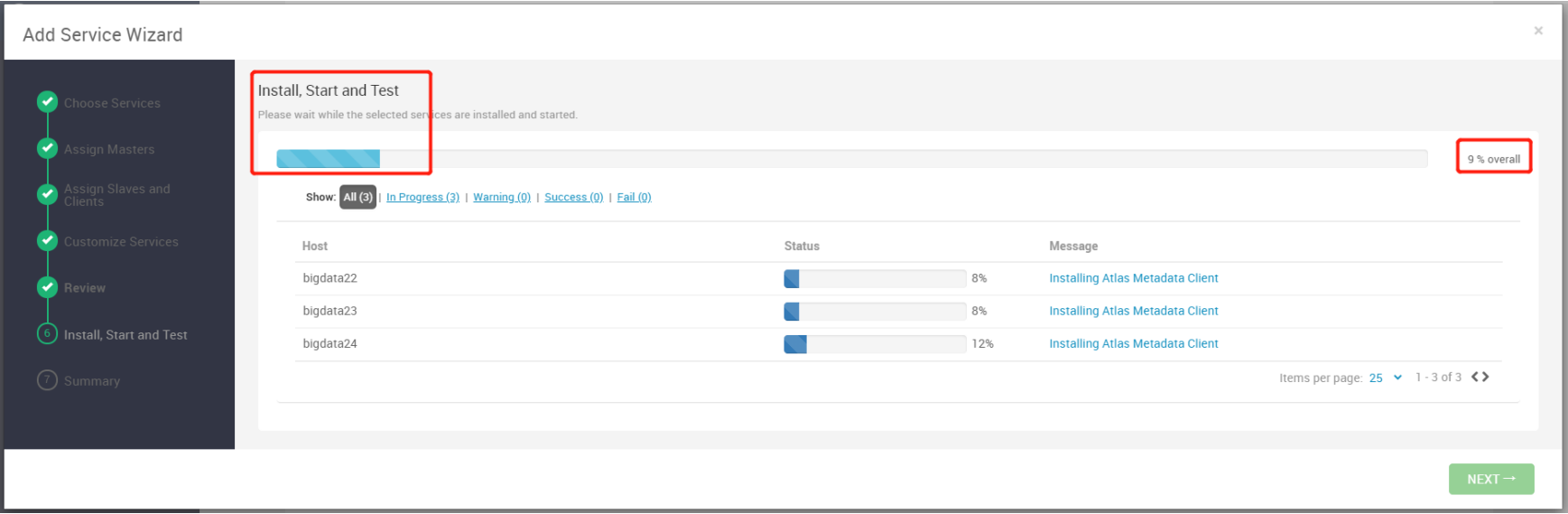
Notebook : bigdata22

← BACK

PRINT

DEPLOY →

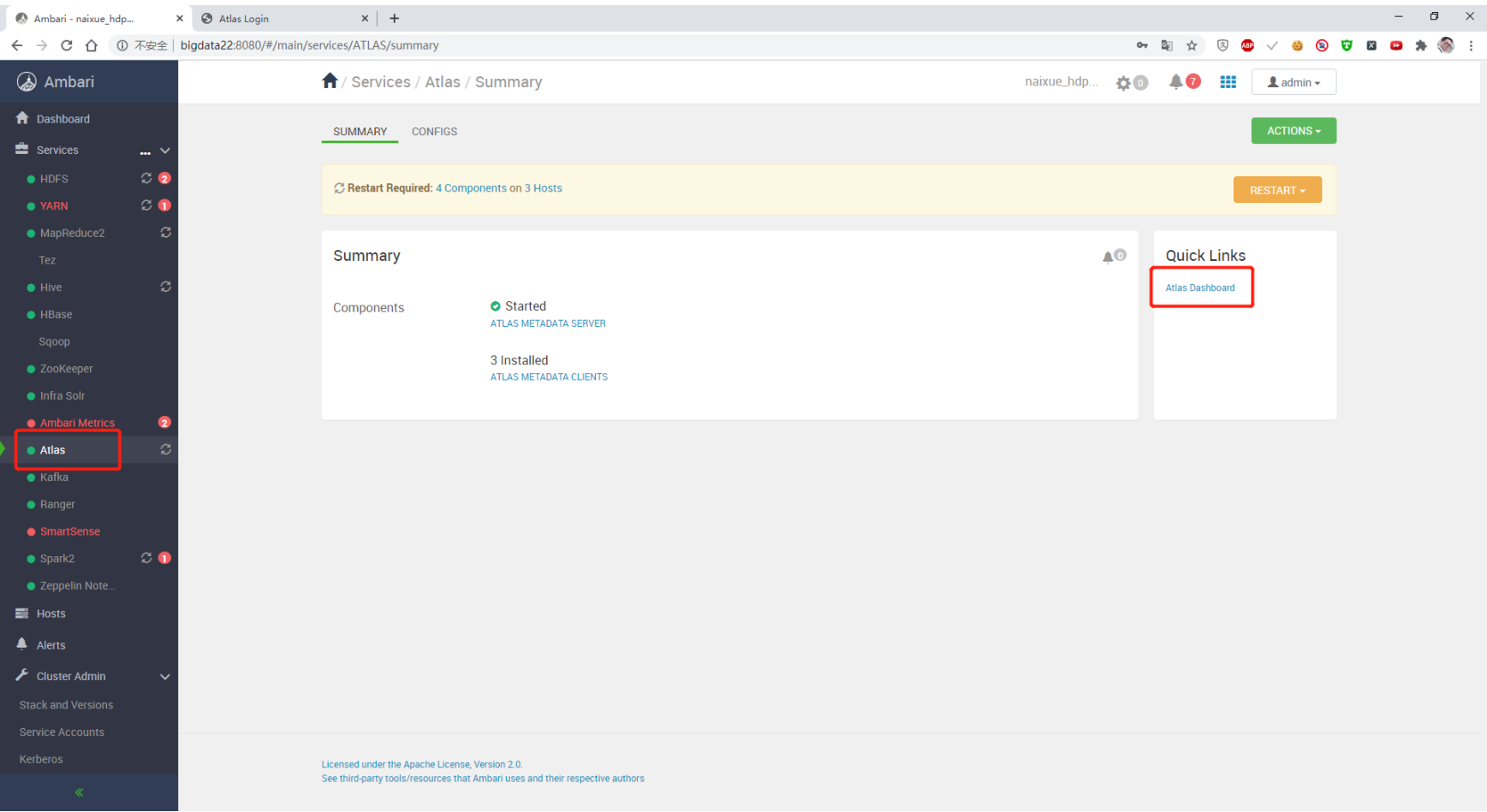
1.8. 等待部署完成



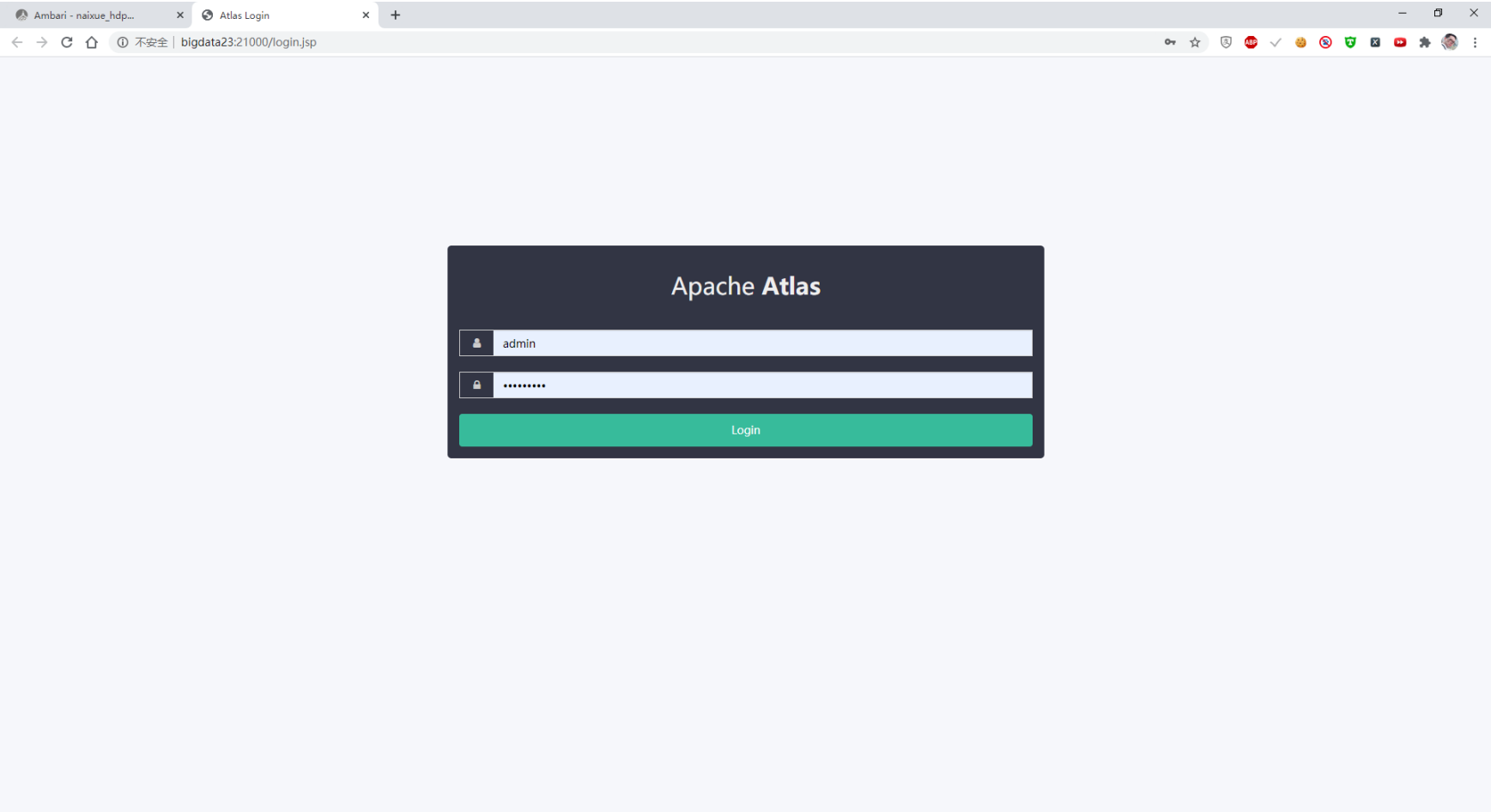
2. 验证

2.1. 验证 Atlas

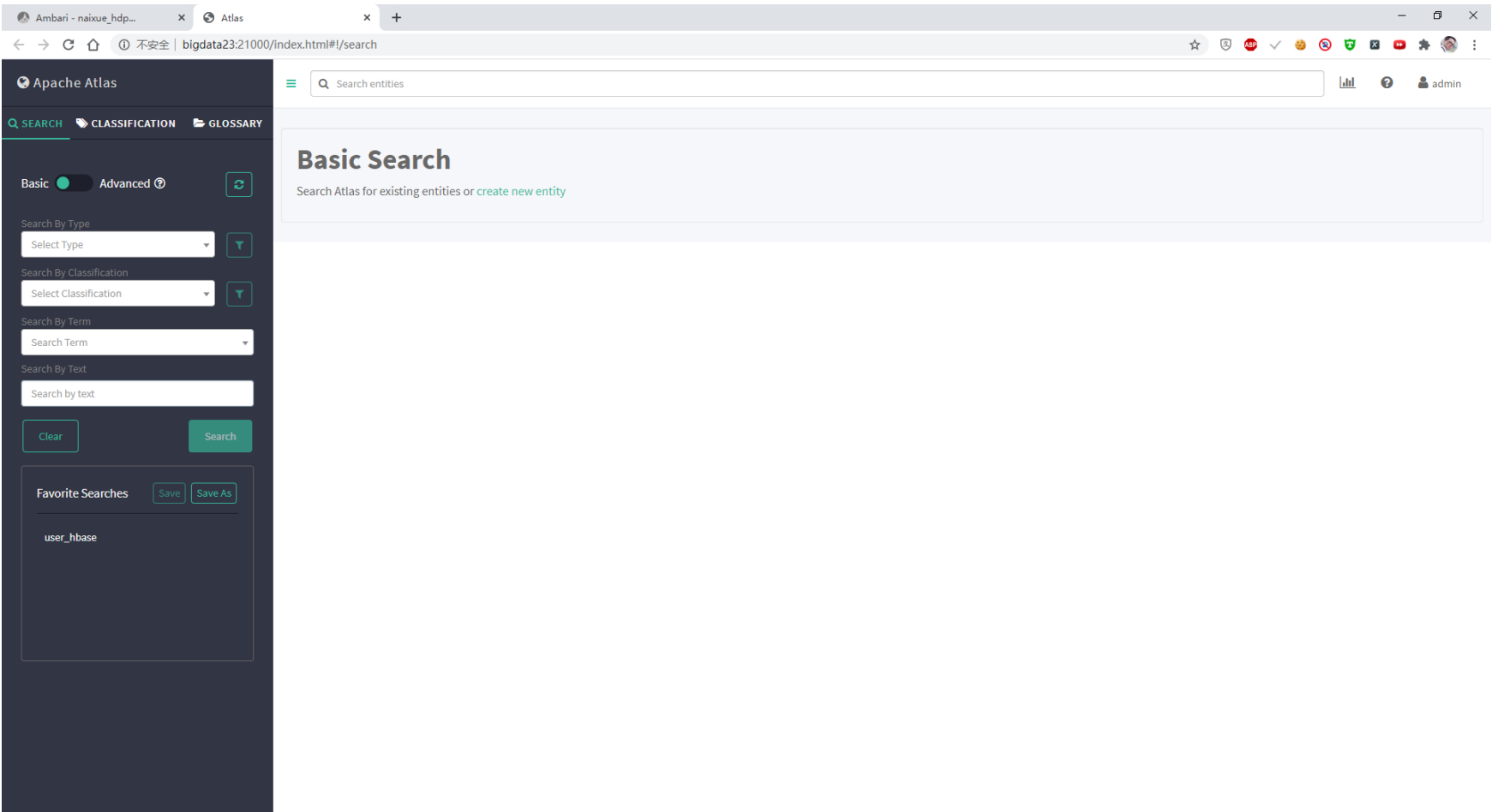
进入 Atlas UI:



登录:



看到主界面:



2.2. 验证 HBase 表是否创建

进入 HBase UI 应该看到 Atlas 创建了两个表：

- ATLAS_ENTITY_AUDIT_EVENTS
- atlas_janus

Atlas 会把元数据和事件存储在这两个表里。

2.3. 验证 Solr 集合是否创建

Atlas 除了把元数据存储 在 HBase，为了实现快速检索，还会把元数据对应的索引存储在 Solr 中，我们登录 Solr 应该可以看到你三个集合已经创建好：

- fulltext_index
- vertex_index
- edge_index

2.4. 验证 Kafka Topic 是否创建

Atlas 会在 Kafka 中创建 2 个 Topic 用于跟外部组件做基于消息的集成：

- ATLAS_HOOK：来自 各个组件的 Hook 的元数据通知事件通过写入到名为 ATLAS_HOOK 的 Kafka topic 发送到 Atlas（消息入口：采集外部元数据）
- ATLAS_ENTITIES：从 Atlas 到其他集成组件（如Ranger）的事件写入到名为 ATLAS_ENTITIES 的 Kafka topic（消息出口：暴露元数据变更事件给外部组件）